



Deductor

Красивая идея
Эффектная реализация...
Реальные проекты...

Аналитическая платформа
для эффективных бизнес-решений

Содержание

Анализ данных	4
Аналитическая платформа	5
Эволюция системы	6
Версия 1	6
Версия 2	7
Версия 3	8
Версия 4-5	9
Тиражирование знаний	10
Состав и назначение модулей	12
Deductor Warehouse	13
Deductor Studio	15
Deductor Viewer	19
Deductor Server/Client	20
Решаемые задачи	21
Истории успеха	23
Сервисная компания “Механика”	23
Коммерческий банк “Стройкредит”	23
Сеть магазинов “Дейли”	24
Образовательная программа	25
Распространение системы	26

Анализ данных

Анализ данных

Анализ данных — это очень широкое понятие. По большому счету любую программу, позволяющую найти нужную информацию, отобразить ее и с помощью полученных данных сделать те или иные выводы, можно назвать аналитическим инструментом. Это действительно так, в процессе анализа приходится использовать различное программное обеспечение, но мы будем рассматривать только специализированные системы, которые ориентированы на решение именно задач анализа.

Почему же мы считаем, что при анализе нужно использовать именно специализированные системы, почему нельзя ограничиться программами общего назначения? Дело в том, что анализ — это не результат каких-то действий, а непрерывный процесс. Конечно, если необходимо раз в месяц построить единственный график, то не стоит ради этого изучать специализированные программы, достаточно воспользоваться стандартными электронными таблицами. Подобный подход вполне применим во многих случаях, но как только потребность в анализе возрастает и его начинают использовать все чаще и чаще, выясняется, что при ручной обработке процесс отнимает очень много времени, а результат оставляет желать лучшего.

Для того чтобы анализировать, вначале нужно как минимум собрать необходимую информацию, что занимает немало времени, ведь обычно данные бессистемно разбросаны по всей организации. Часть информации хранится в учетных системах, что-то в специализированных базах данных, очень много сведений хранится в офисных документах... Это типичная картина практически для всех организаций.

Но даже собрав нужные данные воедино, нельзя сразу их использовать. В подавляющем большинстве случаев первичная информация требует очистки. В данных присутствуют пропуски, аномальные выбросы, противоречия и прочее и прочее. Со всеми этими проблемами нужно как-то бороться, если мы хотим получить действительно качественный результат анализа. Плохое качество исходных данных может не только помешать анализу, но и привести к неверным выводам, что значительно хуже.

Использование специализированного программного обеспечения позволяет не только значительно быстрее решать типовые задачи, возникающие в процессе анализа, например, консолидация данных, очистка, отображение, но и поставить процесс анализа на поток, что гораздо важнее. Именно переход от кустарных или разовых действий по обработке данных к регулярному анализу и вынуждает использовать специализированные аналитические системы.

Может показаться, что для анализа достаточно просто добавить к учетной системе нужные алгоритмы анализа и проблема будет решена, но это не так. Учетные и аналитические системы кардинально отличаются друг от друга.

Сравнительная таблица требований к учетной и аналитической системам:

Характеристика	Учетная система	Аналитическая система
Степень детализации	Первичные документы	Агрегированные данные
Срок хранения	Данные за ограниченный (последний) период	Все накопленные данные за максимально возможный срок
Механизмы обработки	Простейшие математические операции	Сложный математический аппарат
Источники данных	Внутренние операции	Внутренние операции и внешние данные
Характер вычислительной нагрузки	Средняя загрузка с небольшими отклонениями	Периодически возникающая пиковая загрузка

И это только небольшая часть отличий. Требования к учетной и аналитической системам не просто различаются, они противоречат друг другу.

Понимание того, что специализированный инструмент необходим для анализа данных, и привело нас к созданию аналитической платформы Deductor.

Аналитическая платформа

Аналитическая платформа

Возникает вопрос, а почему, собственно, речь идет о платформе, ведь на ее базе нужно еще что-то строить, почему нельзя сразу предоставлять готовые тиражные решения под каждую задачу?³

Дело в том, что анализ очень персонифицирован. Если взять несколько организаций с одной сферой деятельности, то даже в этом случае невозможно модели и правила, полученные для одной компании, без изменений переносить на другую. Алгоритмы анализа могут использоваться одинаковые, но все зависит от того, как их комбинировать и как их настраивать. Все дело в деталях, и именно детали определяют качество анализа.

Например, рассмотрим задачу прогнозирования. Можно ли построить пригодную для всех и не требующую настройки модель прогнозирования спроса на следующий месяц? Конечно! Можно рассчитать прогноз продаж на следующий месяц как среднее от продаж за 2 последних месяца. Все хорошо, пока не начинается оценка качества прогноза. Очень быстро выясняется, что качество неудовлетворительное и что плохой прогноз вообще никому не нужен. Для того чтобы получить хороший, нужно учитывать сезонность для разных товарных групп, особенности работы фирмы, изменения ценовой политики, маркетинговые мероприятия и множество параметров, которые специфичны для каждой компании. Поэтому хороший анализ, в частности, прогноз невозможен без учета особенностей каждой организации, а значит, что для каждой компании модели должны быть свои.

Можно, конечно, использовать заказные аналитические системы, но стоимость подобных программ оказывается высокой, а процесс модификации трудоемким. К тому же нерешенными остаются вопросы качества системы. Заказную систему крайне сложно протестировать во всех возможных режимах так, как это происходит при использовании тиражного программного обеспечения во множестве компаний.

Аналитическая платформа позволяет достичь оптимального баланса между возможностью учесть особенности работы каждой компании, скоростью создания решения и надежностью системы. В платформу уже интегрированы механизмы и инструменты, используемые именно при анализе данных, из-за чего минимизируются сроки создания законченного решения на ее базе. Благодаря наличию способов комбинирования методов анализа можно создать решение, учитывающее специфичные особенности каждой организации. Единое программное обеспечение позволяет гарантировать его качество, т.к. оно тестируется и эксплуатируется во множестве организаций параллельно. Кроме того, отсутствие необходимости программирования и большой набор поставляемых с системой учебно-методических материалов дает возможность практически все работы выполнять самостоятельно и не зависеть от разработчика системы.

Эволюция системы

Версия 1

Эволюция системы

Мы работаем над технологиями анализа данных с 1998 года и все это время развиваем инструменты, которые включены в состав аналитической платформы Deductor. Усовершенствование Deductor шло в зависимости от целей проектов, т.е. в него добавлялись именно те механизмы анализа, которые реально необходимы при решении практических задач. В процессе эволюции у нас менялось представление о том, что такое анализ и каким должно быть аналитическое приложение.

Версия 1

Начав заниматься анализом, мы первым делом реализовали наиболее интересные механизмы, позволяющие по нашему мнению решать актуальные бизнес-задачи. С самого начала мы сделали ставку на самообучающиеся алгоритмы и машинное обучение, как наиболее гибкие и простые в применении механизмы анализа. В результате были реализованы такие технологии, как нейронные сети, самоорганизующиеся карты, деревья решений и т.п.

Самообучающиеся алгоритмы и машинное обучение – общее название классов адаптивных алгоритмов. Их отличительной особенностью является способность самостоятельно находить в больших объемах данных скрытые закономерности. Благодаря этому значительно проще стало обеспечивать их адекватность и актуальность. В случае изменения ситуации достаточно переобучить систему, и модель автоматически учит новые факты. Кроме того, подобные алгоритмы отличаются от классических методов статистического анализа тем, что способны обнаружить сложные закономерности, в частности, нелинейные зависимости.



Deductor 1

Алгоритмы

Эволюция системы

Версия 2

Версия 2

В процессе реализации проектов пришло понимание, что сами по себе алгоритмы не столь принципиально важны, что аналитики оперируют не алгоритмами, а классами задач, например, кластеризации или классификации. Каждый из этих классов задач можно решить множеством способов, в том числе и реализованными нами алгоритмами, но не столь важно как именно. Поэтому в процессе анализа нужно в первую очередь оперировать классами моделей, которых всего 5: кластеризация, регрессия, классификация, ассоциация и последовательность. Эти 5 задач объединяются термином Data Mining.

Задачи, решаемые методами Data Mining:

Data Mining — это процесс обнаружения в «сырых» данных ранее неизвестных нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Классификация — это отнесение объектов к одному из заранее известных классов.

Регрессия — установление зависимости непрерывных выходных переменных от входных значений⁹.

Кластеризация — разбиение объектов на группы (кластеры) таким образом, что объекты внутри кластера были «похожими» друг на друга и отличались от объектов, вошедших в другие кластеры.

Ассоциация — нахождение зависимости, что из события X следует событие Y.

Последовательные шаблоны — установление закономерностей между связанными во времени событиями.



Эволюция системы

Версия 3

Версия 3

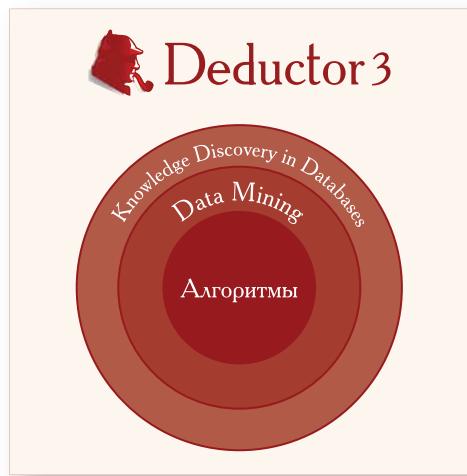
В процессе дальнейшей работы выяснилось, что самая большая проблема при анализе заключается не в методах построения моделей, а в качестве данных. Можно использовать сколь угодно мощные механизмы Data Mining и не получить желаемый результат, т.к. качество реальных данных практически всегда плохое. Если не предпринимать специальных мер для того, чтобы повысить его, например, редактировать аномалии, заполнять пропуски, удалять противоречия и т.п., никакой механизм анализа не позволит получить приемлемый результат, т.е. цикл анализа данных обязательно включает в себя такие этапы, как очистка и предобработка данных. Подобный подход к анализу данных называется Knowledge Discovery in Databases (KDD).

Knowledge Discovery in Databases – это процесс обнаружения знаний в базах данных, включающий этапы выборки, очистки, трансформации, построения моделей и интерпретации результатов. Data Mining является одним из этапов KDD обработки – этапом построения моделей.



Эволюция системы

Версия 3



Версия 4–5

Дальнейшее развитие показало, что и этого недостаточно. Для того чтобы анализ принес реальную отдачу, нужно не только обнаружить закономерности, получить качественный результат, требуется еще сделать так, чтобы результатами анализа могли воспользоваться все, кто заинтересован в нем. Кроме того, необходимо обеспечить возможность удобного просмотра результатов. В принципе необходимо выполнить 4 действия: консолидировать анализируемые данные, формализовать знания, обеспечить прозрачные механизмы работы с полученными знаниями и доставить знания до конечного потребителя. Все вместе это называется процессом тиражирования знаний. Именно для этой задачи знания не только извлекаются из больших объемов данных, но и превращаются в конкурентные преимущества.

Кроме того, необходимо обеспечить максимально простую и безболезненную интеграцию. Аналитическая система не самодостаточна, она должны быть встроена в окружение и обмениваться данными с учетными системами, различными базами данных, хранилищами и прочее. Для реализации «бесшовного сопряжения» в аналитическую систему должны быть встроены различные механизмы интеграции: импорт и экспорт из максимального количества баз, прямая поддержка наиболее популярных систем учета, OLE, удаленный доступ...

Большое количество способов интеграции позволяет выбрать наиболее приемлемый для каждого конкретного случая, иногда достаточно просто выгрузить данные в аналитическую систему и работать с ней, иногда необходимо полностью скрыть механизм принятия решений и взаимодействовать с применением механизмов OLE или специальных API. Подобная реализация дает возможность смотреть на информационную систему, как на единый комплекс, состоящий из модулей, полностью интегрированных между собой.

В результате для реализации механизмов тиражирования знаний в аналитическую платформу были включены сервер, клиент, поддержка OLE-automation и прочие механизмы, позволяющие встраивать Deductor практически в любое окружение.

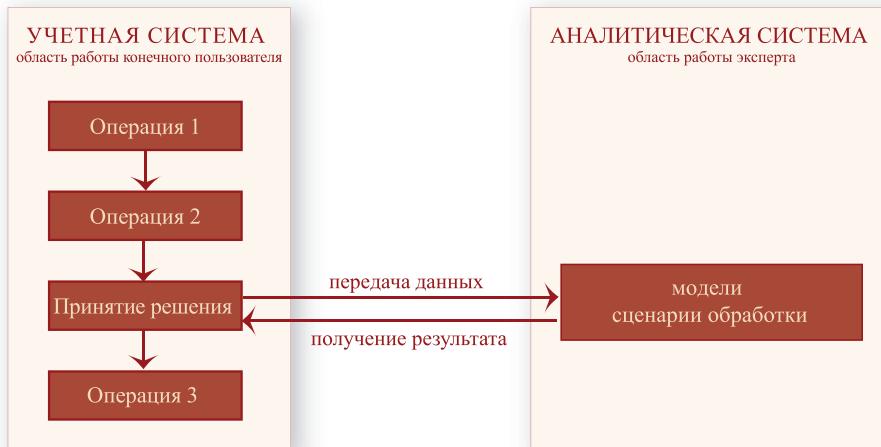


Тиражирование знаний

Тиражирование знаний — это процесс, обеспечивающий сбор данных, формализацию знаний экспертов, визуализацию результатов и их распространение в соответствии с полномочиями. Этот механизм позволяет решить, казалось бы, противоречивую задачу, с одной стороны, повысить качество управления, с другой, снизить требования к персоналу.

Повышение качества управления обеспечивается за счет процесса формализации. Во-первых, любая формализация уже делает работу системы более предсказуемой. Во-вторых, благодаря ей удается использовать опыт лучших сотрудников. Нужно учесть еще один аспект, что речь не идет о простом переносе готовых моделей из одной компании в другую, т.е. в процессе тиражирования можно и нужно учитывать все особенности работы компании, конкурентной среды и прочие факторы.

Снизить требования к персоналу возможно, потому что пользователям не нужно разбираться во всех сложностях анализа. Идеальным можно назвать вариант, когда они даже не замечают существование системы анализа. Реализовать это можно различными способами.



Система может функционировать как аналитическая служба, к которой обращаются из различных сторонних программ для получения рекомендаций, когда необходимо принимать решения. В этом случае обращения к аналитической службе производятся непосредственно из учетной программы и скрыты от конечного пользователя, который просто получает нужные рекомендации и принимает на их основе решения. Работа эксперта заключается в построении моделей и сценариев обработки, которые будут использоваться при принятии решений. Возможен и другой вариант построения системы. Данные, имеющиеся в различных источниках, консолидируются, строятся сценарии, позволяющие извлекать из консолидированных данных правила, закономерности, прогнозы. Результаты обработки отображаются при помощи специализированного приложения для просмотра отчетов. Как и в первом варианте, конечный пользователь не задумывается, каким образом получены эти прогнозы, т.к. сделать это очень просто — надо выбрать один из отчетов. Все необходимые для получения результата действия система должна произвести самостоятельно.

Таким образом, можно отделить работу аналитиков от работы конечных пользователей. Если потребуется, то можно дать возможность конечному пользователю глубже понять, каким образом получено решение, но в большинстве случаев у сотрудников нет ни желания, ни возможности в чем-либо разбираться. Им нужен готовый результат без головной боли.

В результате благодаря тиражированию знаний можно не просто их извлечь из данных, но и превратить в конкурентные преимущества, что и является основной задачей любой аналитической системы.

Тиражирование знаний



Для того чтобы система могла использоваться как инструмент тиражирования знаний, необходима реализация в ней следующего функционала:

Встроенные механизмы консолидации данных.

При получении результатов анализа пользователь не должен беспокоиться по поводу доступности данных, их корректности и качества. Вся информация должна уже существовать в пригодном для анализа виде. Поэтому наличие готовых инструментов сбора, консолидации и хранения данных является критически важной частью системы.

Механизмы очистки, предобработки и построения моделей.

Необходим полноценный KDD инструмент для формализации знаний. Все используемые механизмы должны иметь прозрачный, требующий минимальных усилий способ комбинирования действий по обработке данных. Критически важным является реализация такого способа комбинирования методов, который позволил бы обеспечить автоматическую потоковую обработку данных, в противном случае можно будет получить разовый результат, но невозможно — реализовать массовую обработку данных.

Механизмы отображения результатов.

Результаты анализа должны быть представлены в наиболее доступном для понимания виде и скрывать внутреннюю сложность обработки. Эти механизмы должны быть ориентированы на конечного пользователя, а не на эксперта в области математики или статистики.

Тиражирование знаний

Механизмы интеграции.

Результаты анализа должны выйти за пределы аналитической системы, необходимо обеспечить возможность их использования непосредственно в бизнес-процессе. Невозможно выдвигать какие-то специфические требования к используемым в компании бизнес-приложениям, которых в организации может быть множество и которые взаимодействуют различными способами. Бизнес-приложения могут быть распределенными, базироваться на различных СУБД, использовать специфические протоколы обмена данными. Аналитическая система должна интегрироваться с любым окружением. Для этого нужно иметь максимально гибкие возможности интеграции: поддержку различных баз данных и бизнес-приложений, возможность удаленной обработки, возможность взаимодействия с использованием API, OLE и прочее.

Механизмы разграничения доступа.

Это необходимо для того, чтобы отделить работу эксперта от работы конечного пользователя. От конечного пользователя нужно скрыть сложности обработки, кроме того, он не должен иметь возможность править существующие модели, т.к. правка моделей без соответствующей квалификации может принести только вред. Помимо этого, нужно иметь возможность защитить построенные модели и сценарии от несанкционированного доступа и распространения.

Весь описанный функционал был реализован в аналитической платформе Deductor, основной задачей которой является тиражирование знаний. В программу встроено множество алгоритмов анализа, различные способы визуализации, механизмы консолидации и т.п., но все они являются всего лишь инструментами для решения главной задачи — тиражирования знаний.

Ценность Deductor не в том или ином реализованном механизме анализа, а в интегриированном подходе, позволяющем поставить процесс анализа данных на поток.

Состав и назначение модулей

Deductor состоит из 5-ти интегрированных модулей:

- **Warehouse** — хранилище данных
- **Studio** — рабочее место аналитика
- **Viewer** — рабочее место конечного пользователя
- **Server** — сервер для удаленной аналитической обработки
- **Client** — библиотека для работы с Deductor Server

Состав и назначение модулей Deductor Warehouse

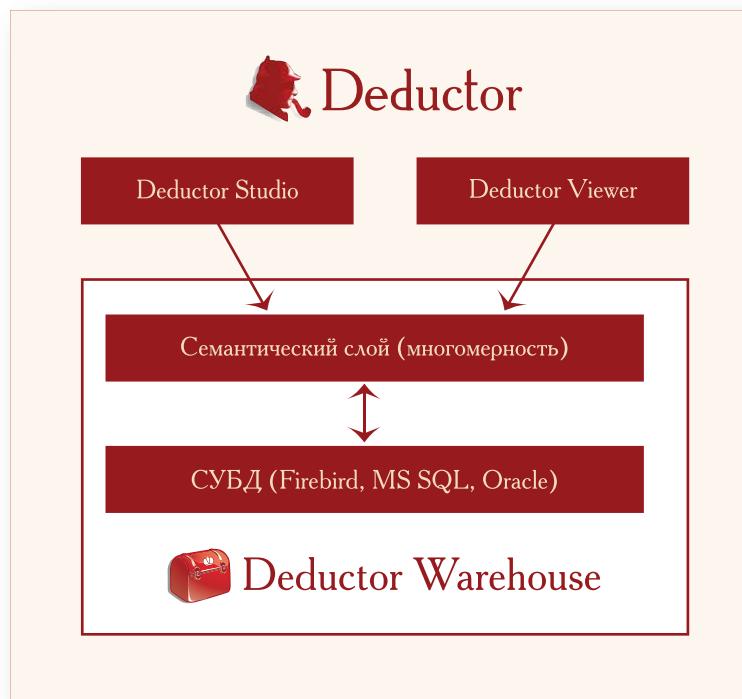
Deductor Warehouse

Deductor Warehouse — многомерное хранилище данных, предназначенное для решения задачи консолидации информации. Использование единого хранилища позволяет обеспечить простой и прозрачный доступ к данным, контроль целостности и непротиворечивости информации, высокую скорость обработки. Благодаря глубокой степени интеграции любую информацию из хранилища данных можно получить в приложениях Deductor с минимальными усилиями.

Хранилище данных ориентировано именно на аналитическую обработку, поэтому включает в себя все, что необходимо для комфортной работы при анализе. Оно содержит интегрированный семантический слой, т.е. механизм, автоматически преобразовывающий бизнес-термины в операции с базой данных и обратно. Благодаря наличию семантического слоя пользователь оперирует такими бизнес-понятиями, как «клиент», «товар», «прибыль», а система автоматически выполняет необходимые действия с базой данных и предоставляет пользователю нужную информацию.

Применение хранилища данных позволяет не быть привязанным к учетной системе, хранить данные не только за последний период, а за весь необходимый для анализа срок, консолидировать информацию из разнородных источников. Использование специализированных методов хранения и извлечения данных значительно увеличивает скорость получения информации. Хотя наличие единого источника данных не является обязательным условием работы аналитической системы, практически всегда ее создание начинается с построения хранилища данных.

Схема доступа к хранилищу данных:



Deductor Warehouse поддерживает прозрачную работу с тремя СУБД: Firebird, MS SQL и Oracle. вне зависимости от используемой СУБД работа с хранилищем происходит совершенно одинаково с использованием единого унифицированного механизма доступа.

Поддержка нескольких СУБД в качестве платформы хранилищ данных позволяет в каждом конкретном случае применять наиболее пригодную для данного случая базу данных. Это дает возможность минимизировать стоимость приобретаемых лицензий, комбинируя как коммерческое (Oracle, MS SQL), так и бесплатное (Firebird) программное обеспечение. В зависимости от особенностей функционирования организации возможно построение системы на базе единого хранилища данных, набора хранилищ, комбинации хранилищ, витрин данных и прочее.

Состав и назначение модулей Deductor Warehouse

Вот несколько вариантов реализации концепции хранилища данных с использованием Deductor Warehouse:

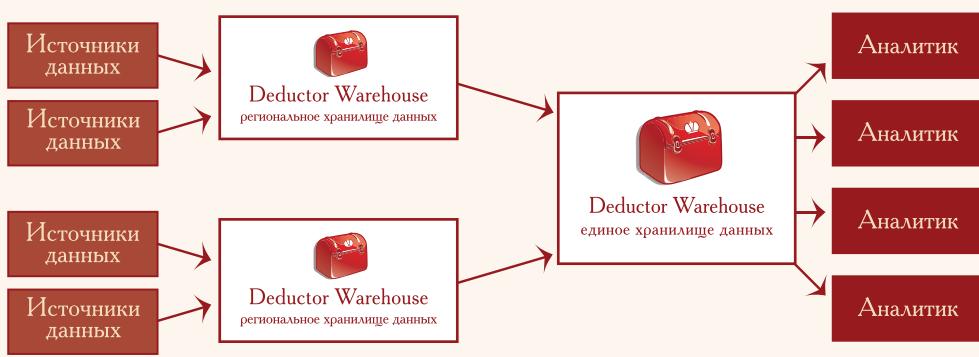
Единое хранилище данных



Единое хранилище данных и витрины данных



Региональные базы с интегрированным хранилищем данных



Кроме того, в Deductor реализована поддержка концепции виртуальных хранилищ данных — Virtual Warehouse. Виртуальное хранилище данных обеспечивает прозрачный для аналитика доступ к сведениям, хранящимся в любых реляционных СУБД. Взаимодействие с Virtual Warehouse происходит аналогично работе с традиционным хранилищем данных. Аналитик оперирует бизнес-понятиями, заданными в семантическом слое, и от него скрыты все сложности выборки данных, как и в случае с Deductor Warehouse. Пользователь задает при помощи простого Мастера, какая информация его интересует, а система автоматически трансформирует их в запросы к базе данных.

Таким образом, эмулируется работа хранилища данных, т.е. данные реально не перегружаются в специализированную систему, все операции производятся «на лету». Virtual Warehouse позволяет представить информацию, хранящуюся в реляционных базах данных, в удобном для аналитика многомерном виде.

Состав и назначение модулей Deductor Studio

Deductor Studio

Deductor Studio — рабочее место аналитика. В этом приложении осуществляется формализация знаний эксперта. Программа включает все необходимые для анализа инструменты обработки: механизмы импорта данных из разнородных источников, методы очистки и предобработки, алгоритмы построения моделей и механизмы экспорта данных.

Схема работы Deductor Studio:



Все действия по анализу данных сводятся всего к 4-м операциям:

1. Импорт данных.

В процессе импорта данные получаются из источника и загружаются специальным образом в программу. В дальнейшем с ними можно производить любые доступные операции, работа со всякой импортированной таблицей происходит одинаково. Поддерживается импорт из наиболее распространенных СУБД (Oracle, MS SQL, MySQL, Interbase...), стандартных файлов обмена данными (dbf, txt, csv...), офисных приложений (MS Excel, MS Access...), бизнес-программ (1С v7, v8...). Кроме того, в программу встроен механизм импорта с применением стандартов доступа к данным ODBC и ADO.

2. Обработка данных.

Обработкой называется любое действие над данными, приводящее к их преобразованию, например, очистка данных либо построение моделей. Ее результатом является набор данных, который можно опять обработать каким-либо способом. Благодаря этому обеспечивается возможность построения сценариев обработки, т.е. последовательных операций над данными, приводящих к нужному результату. Поддерживается широкий набор механизмов обработки: методы очистки (заполнение пропусков, редактирование аномалий, фильтрация...), инструменты предобработки (квантование, группировки, сортировки...), методы построения моделей (нейронные сети, самоорганизующиеся карты, деревья решений...).

Состав и назначение модулей Deductor Studio

3. Визуализация.

Полученные результаты можно просмотреть различными способами, начиная от простых таблиц и диаграмм до многомерных кубов и специализированных визуализаторов. Система построена таким образом, что самостоятельно определяет возможные способы визуализации и предлагает наиболее удобные способы отображения данных для каждого случая.

4. Экспорт данных.

Результаты обработки могут быть выгружены во множество приемников данных. Таким образом, обработанная и проанализированная информация выходит за пределы аналитической платформы, попадает в бизнес-приложения, офисные программы и прочее.

В Studio реализованы самые современные самообучающиеся алгоритмы анализа, но не это главное. Важно то, что любые способы обработки можно комбинировать произвольным образом. Именно за счет комбинирования различных подходов и методов анализа можно получить действительно качественные результаты.

Анализ данных в Deductor Studio базируется на построении сценариев обработки. Подобный подход к анализу интуитивно понятен большинству аналитиков, т.к. анализируя данные, вне зависимости от инструментария они фактически готовят сценарии обработки.

Как выглядит типовой сценарий, например, при построении прогнозов? Сначала аналитик загружает анализируемые данные в Excel, что и является операцией импорта. Потом он проверяет данные на наличие ошибок и исправляет их, например, продажи с нулевой суммой или возврат товара поставщику, что есть операция очистки. Далее он группирует данные для получения итоговой информации по месячным продажам определенного товара — это операции трансформации. Потом аналитик пытается подобрать полином или другую формулу, которые объясняли, почему раньше продажи вели себя именно так, — это этап построения модели. Далее он применяет построенное правило для получения прогноза на следующий месяц, что, собственно, является прогнозированием. И последний этап анализа — отправка результатов прогноза заинтересованному лицу, этот процесс называется экспортом. Работая с Deductor, аналитик строит сценарий, который очень схожий с тем, что описан. Именно поэтому для аналитика этот процесс интуитивно понятен.

Ценность Deductor Studio еще и в том, что сценарная методология анализа в него изначально заложена, система не позволяет работать по-другому. На первый взгляд, данная особенность программы может показаться неудобной, но это не так. Вернемся к примеру с прогнозированием в Excel и представим, каким образом аналитик, скорее всего, будет работать, если встретит в данных ошибки. Наверняка он выделит неверные записи, удалит их и продолжит работу. Его действия будут совершенно естественны, т.к. именно таким образом аналитик сможет быстрее всего получить результат. Он прав, но только при условии, что анализ выполняется один раз. Если в следующий раз аналитик столкнется с такой же проблемой, он будет должен снова выполнить те же самые действия. То, что сделал аналитик в первый раз, т.е. исправил вручную, не тиражируемо, это нельзя поставить на поток. При выполнении того же самого в Deductor Studio все будет совершенно иначе.

Deductor Studio не имеет механизмов ввода и ручной правки данных. Подобное ограничение на функциональность наложено осознано. В случае, если аналитик, получив данные, обнаружит в них, например, ошибки, он должен будет описать правило работы с такими данными. В примере, что был дан выше, он должен будет отфильтровать данные о продажах с нулевой суммой. Это нужно сделать обязательно, т.к. вручную в Deductor Studio при всем желании удалить непригодные записи невозможно. То, что он сформулирует, автоматически станет частью сценария. Такая работа требует чуть больше усилий и времени, чем простое удаление данных из электронной таблицы, но подобный сценарий обработки тиражируем. При появлении новых данных не нужно опять искать некорректные записи, т.к. правила их обработки уже имеются в сценарии и очистка данных может быть выполнена автоматически. Именно эта особенность и позволяет говорить о Deductor как об инструменте тиражирования знаний.

Состав и назначение модулей Deductor Studio

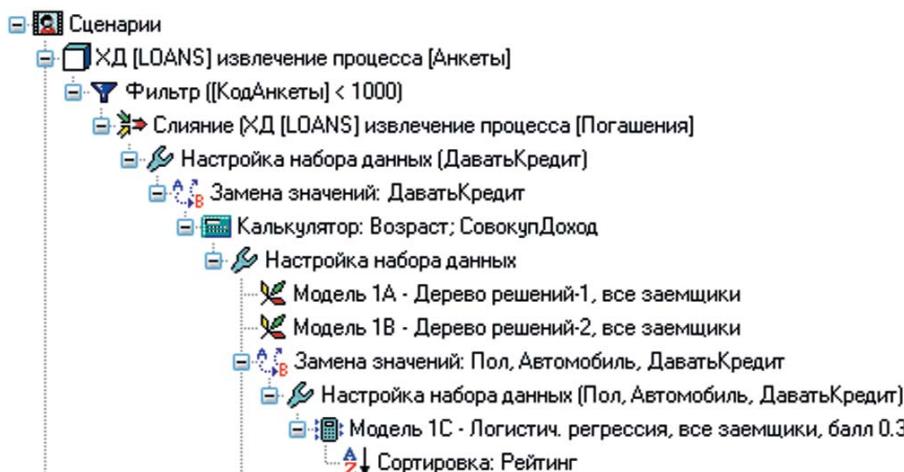
Подготовленные сценарии можно выполнять автоматически в пакетном режиме, например, ночью. В результате аналитик единожды готовит сценарии, и через них в последствии просто «прогоняются» новые данные. Подобный механизм работает не только для построения прогнозов, но и для автоматической загрузки данных в хранилище Deductor Warehouse, например. Другими словами, Deductor можно использовать не только как систему анализа, но и как ETL средство — инструмент извлечения, трансформации и загрузки данных. А так как аналитик имеет возможность произвести загрузку данных после любой последовательности обработки, можно не просто загружать данные в хранилище, а собрать их из разрозненных источников, объединить, очистить, обработать любым способом, привести к нужному виду и загрузить в хранилище более ценные и практически полезные сведения.



Анализ данных — это не разовая операция, поэтому нужно обеспечить возможность полноценной поддержки этого процесса. Периодически приходится модифицировать сценарии, и необходимо сделать эту операцию удобной. Особенности человеческой памяти таковы, что даже создатель сценария спустя некоторое время не может в точности вспомнить логику ее построения. Нужен способ документирования процесса формализации. В Deductor это учтено, сценарии отображаются в виде дерева с иконками и пояснительным текстом. Взглянув на это дерево, можно без труда проследить логику сценария и понять особенности его реализации. Это помогает не только модифицировать сценарии, но и передавать их другому аналитику, который также просто сможет «прочесть» ход мысли аналитика, создавшего сценарий.

Анализ не ограничивается только обработкой данных, не менее значимой является визуализация данных. Известно, что через органы зрения человек получает подавляющую часть информации и, конечно, нужно по максимуму использовать эту особенность человеческого восприятия. Хорошая визуализация способна серьезно помочь при анализе данных, кроме того, очень много идей рождается именно при просмотре данных.

Способов визуализации существует огромное множество, она не ограничивается таблицами и диаграммами, информацию можно просмотреть в виде кросс-таблиц, гистограмм, карт, деревьев и т.п. Чем больше механизм визуализации учитывает особенности обработки данных, тем больше от него отдачи. Например, если мы просматриваем результаты прогнозирования в виде графика, желательно выделить цветом или каким другим образом прогнозные и исторические значения, отметить данные выходящие за пределы допустимого интервала. Все это, не требуя никаких усилий со стороны аналитика, делает визуализацию намного более ценной и упрощает работу пользователя.



Состав и назначение модулей Deductor Studio

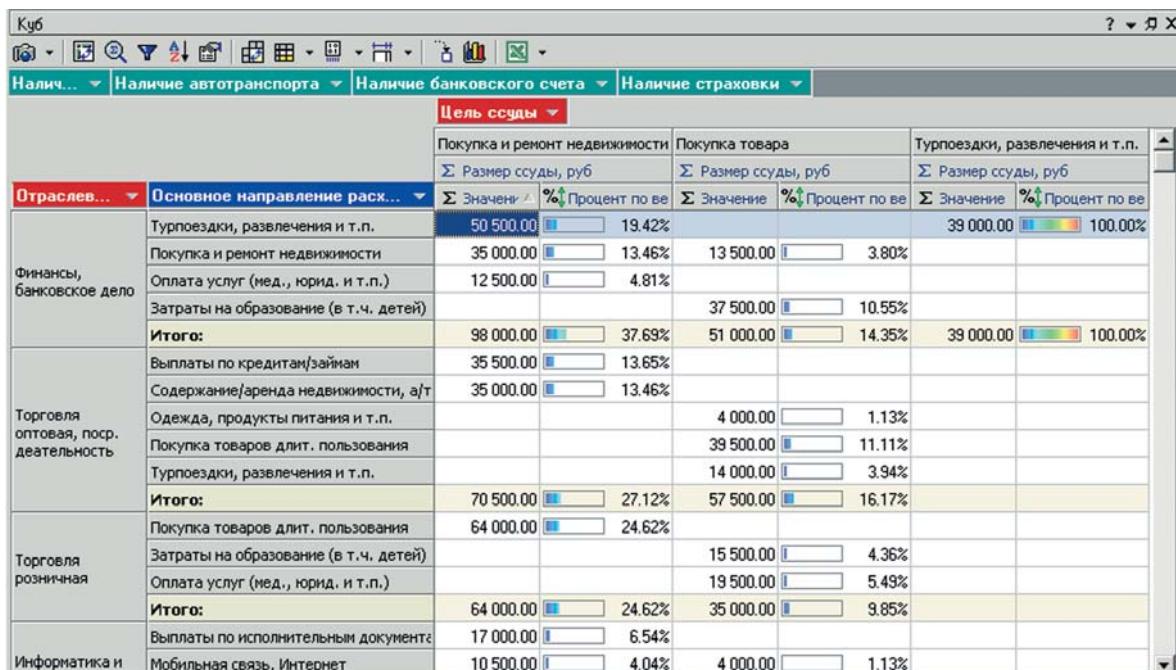
Эти особенности визуализации учтены в Deductor Studio. В системе имеется множество удобных способов отображения данных. Программа самостоятельно анализирует способы обработки, особенности набора данных, на которых производился анализ и автоматически предлагает возможные способы визуализации.

Среди множества механизмов визуализации, встроенных в Deductor Studio, имеется и мощный Online Analytical Processing (OLAP) модуль. OLAP — один из наиболее популярных способов отображения табличных данных. Данные в этом случае могут отображаться в виде кросс-таблиц или кросс-диаграмм. Кросс-таблицы удобны тем, что большая часть операций манипулирования данных выполняется «на лету». Одним щелчком мыши можно данные сгруппировать произвольным образом, отфильтровать, отсортировать, переставить столбцы/строки и произвести множество других операций. Deductor Studio позволяет при помощи этого механизма визуализации просмотреть любые данные, т.е. не только саму исходную информацию, но и результаты любой обработки. Deductor Studio — это инструмент аналитика, а он является ключевым лицом в процессе анализа данных, именно его знания формализуются и тиражируются, но многие пользователи не являются аналитиками, разбираясь во всех трудностях обработки им сложно, для них нужен более простой и понятный способ получения требуемой информации.

Для скрытия от конечных пользователей всех особенностей обработки данных используется большое число механизмов: можно взаимодействовать с Deductor Studio при помощи механизма OLE-Automation, воспользоваться удаленной обработкой в Deductor Server, в автоматическом режиме выгружать и загружать данные в различные системы, но имеется возможность получить результат еще легче, приложив минимум усилий.

Это можно сделать благодаря механизму получения готовых отчетов. В Deductor Studio имеется панель отчетов, внешне напоминающая проводник в Windows. На этой панели аналитик формирует иерархическую структуру папок и в определенные папки выносит ссылки на интересующие пользователей узлы сценария. Например, пользователя интересует прогноз продаж. Прогноз продаж — это результат выполнения целой цепочки действий, но в папке на панель отчетов выносится только последний узел — получение, собственно, самого прогноза. Когда пользователь выберет интересующий его отчет на панели, все действия по извлечению данных и выполнению последовательности действий, необходимых для получения результата, будут произведены автоматически, конечный пользователь просто получит отчет о прогнозируемых продажах.

Таким образом, обеспечивается разделение обязанностей между аналитиками и конечными пользователями. Для последних в Deductor имеется специальное приложение — Deductor Viewer.



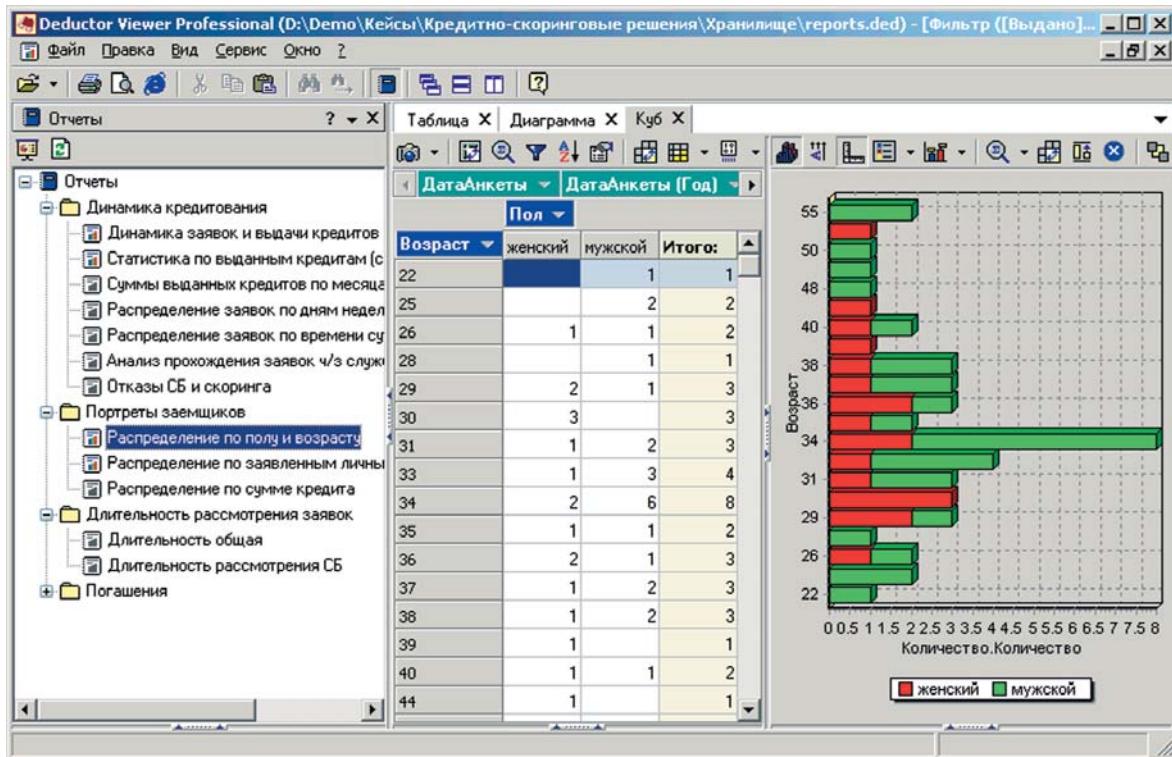
Состав и назначение модулей

Deductor Viewer

Deductor Viewer

Deductor Viewer — это рабочее место конечного пользователя. В нем отсутствуют механизмы построения сценариев, настройки источников данных и прочие сложности. Работа с программой упрощена до предела: пользователь видит настроенную аналитиком панель отчетов, выбирает интересующий отчет, программа автоматически выполняет все необходимые действия и конечный пользователь получает результат.

Полученный результат можно просмотреть разными способами, изменить настройки отображения, перенести в офисные приложения, распечатать, но не более того. Пользователь Viewer не может изменить сами результаты. Подобные ограничения могут показаться слишком строгими, но именно это и нужно большинству пользователей: выбрал отчет — получил ответ.

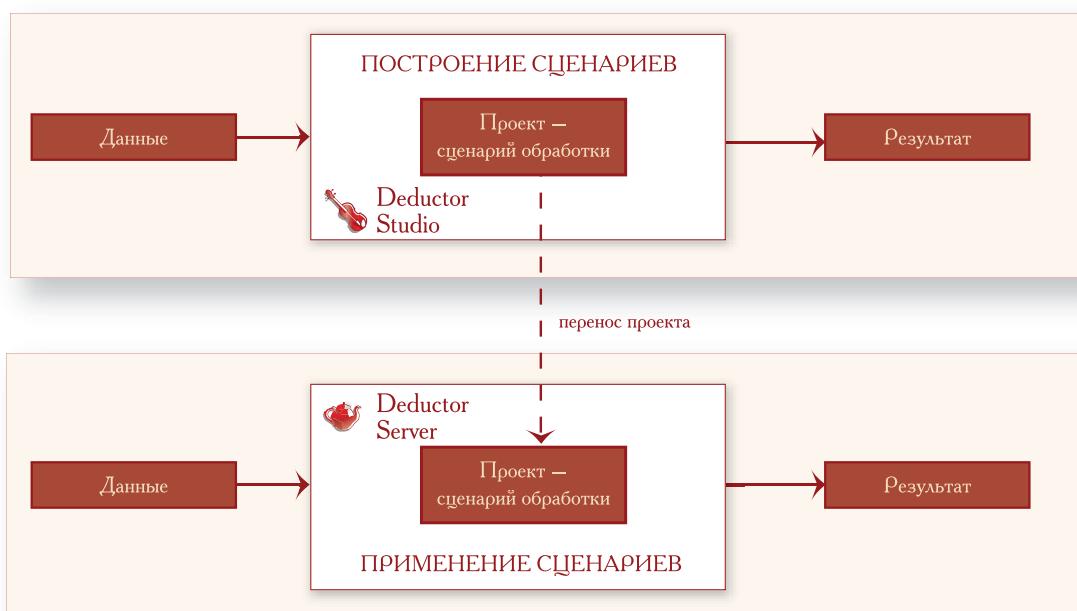


Состав и назначение модулей Deductor Server/Client

Deductor Server/Client

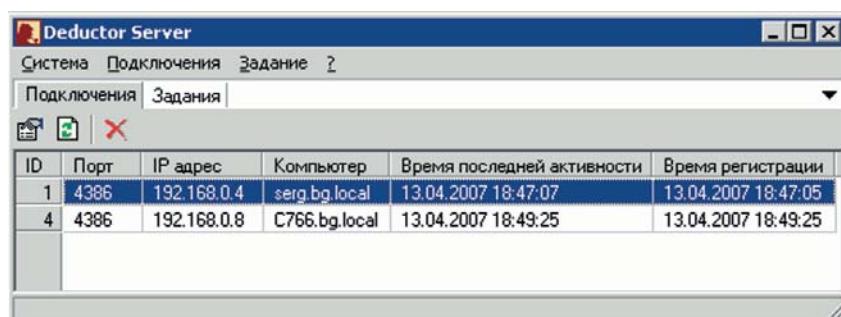
Процесс построения сценариев является интерактивным, т.к. необходимо строить модели, проверять результаты, анализировать качество... Аналитик эти операции выполняет помногу раз, пока не будет спроектирован приемлемый сценарий и построены качественные модели. Для подобной работы аналитику требуется специализированное рабочее место, каковым является Deductor Studio, но после построения сценария можно перейти к автоматизированной обработке данных. Необходимо, чтобы вновь поступающие данные автоматически «прогонялись» через разработанные аналитиком модели уже без участия эксперта.

Для выполнения этой функции в состав платформы Deductor входит сервер. С его помощью можно автоматически обрабатывать сценарии. Deductor Server функционирует в виде Windows-службы, к которой можно обращаться удаленно при помощи специального клиента - Deductor Client. Управлять выполнением сценарием можно как из локальной сети, так и через Интернет.



Применение сервера позволяет обеспечить удаленную обработку, оптимизировать расход оперативной памяти, повысить скорость обработки, обеспечить прозрачную интеграцию со сторонними приложениями. Deductor Server специальным образом кэширует данные, что позволяет их использовать без обращения к дисковой системе. Это значительно повышает скорость аналитической обработки. При применении многопроцессорного оборудования обеспечивается увеличение скорости благодаря встроенной многопоточной обработке и механизмам балансировки нагрузки.

Использование Deductor Server значительно упрощает создание полноценной корпоративной аналитической системы, его применение позволяет воспользоваться всеми преимуществами трехзвенной архитектуры, оптимально используя возможности серверной аналитической обработки.



Решаемые задачи

Решаемые задачи

Идеи, реализованные в Deductor, с успехом используются для решения самых разнообразных аналитических задач. Фактически они могут применяться для обработки любой табличной информации, сфера деятельности организации не имеет особого значения.

Deductor применяется для анализа данных в оптовой и розничной торговле, в банках и страховых компаниях, на промышленных предприятиях и в органах государственного управления.

Вот только небольшой список, решаемых при помощи аналитической платформы, задач:

Оптовая торговля:

- Аналитическая отчетность;
- Прогнозирование спроса;
- Контроль товаров с ограниченным сроком годности;
- Сегментация клиентской базы;
- Оценка эффективности маркетинговых действий;
- Оптимизация закупок;
- Оптимизация продуктовой линейки.

Розничная торговля:

- Консолидации данных;
- Аналитическая отчетность;
- Прогнозирование спроса;
- Оценка эффективности рекламной компании;
- Оценка эффективности работы подразделений;
- Оптимизация ценовой политики;
- Оптимизация размещения товаров;
- Оптимизация системы скидок и бонусов.

Банки:

- Оценка кредитоспособности физических лиц;
- Оценка кредитоспособности юридических лиц;
- Анализ клиентской базы;
- Анализ структуры доходов и расходов;
- Управленческая отчетность;
- Консолидация данных ;
- Прогнозирование финансовых поступлений.

Решаемые задачи

Промышленные предприятия:

- Анализ движения денежных средств;
- Контроль качества;
- Обнаружение дефектов;
- Анализ потребительских характеристик производимых товаров;
- Прогнозирование спроса;
- Анализ качества сырья.

Deductor продемонстрировал свою эффективность во множестве реальных проектов

Сфера деятельности	Организация	Применение
Банк	Русь-банк	Оценка кредитоспособности физических лиц
Банк	Севергазбанк	Оценка кредитоспособности юридических лиц
Банк	Стройкредит	Оценка кредитоспособности физических лиц, аналитическая отчетность
Оптовая торговля	МДМ-комплект	Консолидация данных, аналитическая отчетность, прогнозирование
Оптовая торговля	Медарго	Прогнозирование спроса, оптимизация закупок, контроль сроков годности, анализ отклонений
Оптовая торговля	Сонекс	Консолидация данных, аналитическая отчетность, прогнозирование
Производство	Элбург	Прогнозирование спроса, оптимизация продаж, аналитическая отчетность
Производство	ЗАО «Веда»	Консолидация, отчетность, анализ клиентской базы
Промышленность	Надымгазпром	Диагностика трубопроводов
Промышленность	Световые технологии	Консолидация данных, финансовая отчетность, управленческая отчетность, прогнозирование
Промышленность	Шатурская мебель	Анализ производства, анализ маркетинговой информации
Розничная торговля	Дейли фудс	Консолидация, аналитическая отчетность, оптимизация ценовой политики
Розничная торговля	Комус	Прогнозирование, анализ клиентской базы
Сервис	Механика	Консолидация данных, прогнозирование спроса, оптимизация закупок
Транспорт	Новая перевозочная компания	Анализ транспортных потоков

Для того чтобы стало яснее, кто и как использует Deductor, приведем описание некоторых реализованных проектов.

Истории успеха

Сервисная компания "Механика"

Истории успеха

Сервисная компания "Механика"

Сервисная компания «Механика» работает на российском автомобильном рынке с 1991 года. Основным видом деятельности компании является капитальный ремонт двигателей и поставка запасных частей.

Для стабильной работы компании необходимо постоянное наличие на складе нужных деталей и запчастей. В случае небольшой номенклатуры данная проблема решается относительно просто, но в «Механике» номенклатурных позиций несколько десятков тысяч. Такой большой ассортимент делает невозможной ручную обработку данных, поэтому было решено перейти на автоматизированные способы прогнозирования и формирования заказа.

Система, построенная на платформе Deductor, позволила решать поставленную задачу. Первоначально было создано хранилище, в которое была загружена необходимая для анализа информация из учетной системы компании. На ее основе были построены модели прогнозирования и разработаны сценарии их применения. Результаты анализа передавались обратно в учетную систему для формирования заказа.



Запущенная система позволила поднять качество анализа на совершенно новый уровень. Стало возможно оперативное управление закупками. Прогнозы автоматически формируются каждую неделю, т.к. не возникает сложностей с обработкой большой номенклатуры. Значительная часть операций выполняется автоматически в ночное время, в результате снижается нагрузка как на оборудование в наиболее напряженные часы работы, так и на пользователей, т.к. они каждое утро имеют нужные данные, представленные в удобном виде.

Коммерческий банк "Стройкредит"

Коммерческий банк «Стройкредит» основан в 1988 году, в 2005 году вышел на рынок потребительского кредитования.

Для работы на этом рынке была необходима скоринговая система, позволяющая оценивать риски при выдаче кредитов. Так как речь шла о массовом потребительском кредитовании относительно небольших сумм, нужна была система, дающая возможность значительно повысить скорость принятия решений, не увеличивая риски.

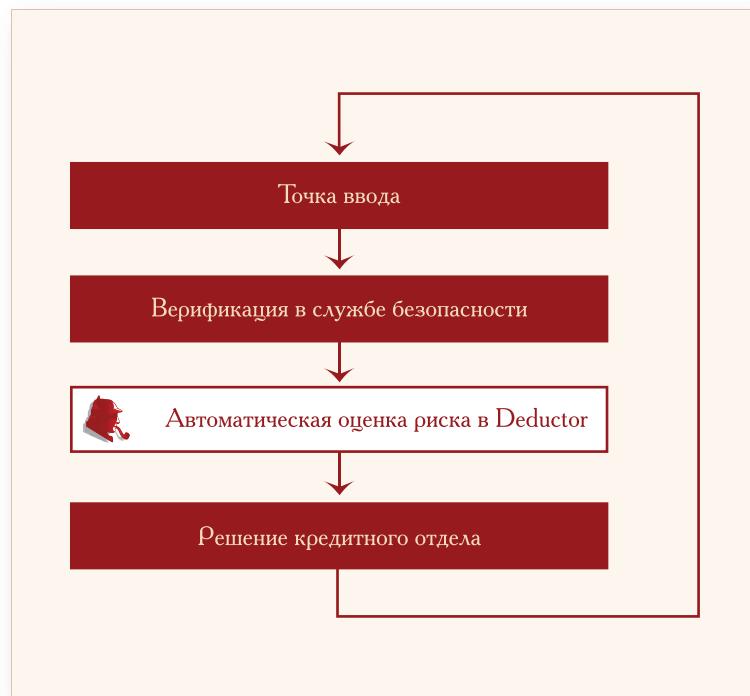
Такая система была разработана на базе аналитической платформы Deductor и интегрирована с существующей в организации автоматизированной банковской системой.

Работа пользователей с системой строится на базе веб-технологий. В месте выдачи кредита заполняется веб-форма заявки на получение кредита. Эта информация попадает в центральный офис банка, в котором выполняются все необходимые операции, автоматически в фоновом режиме оцениваются риски при помощи реализованной в Deductor скоринговой модели и на основе этой оценки принимается решение о выдаче кредита.

Запущенная система позволила увеличить скорость обработки заявок, не повышая рисков и реализовать действительно потоковую обработку данных. Использование же аддитивных и самообучающихся алгоритмов анализа обеспечивает возможность простого внесения изменений в скоринговую модель для учета самых разнообразных факторов.

Истории успеха

Сеть магазинов "Дейли"



Сеть магазинов "Дейли"

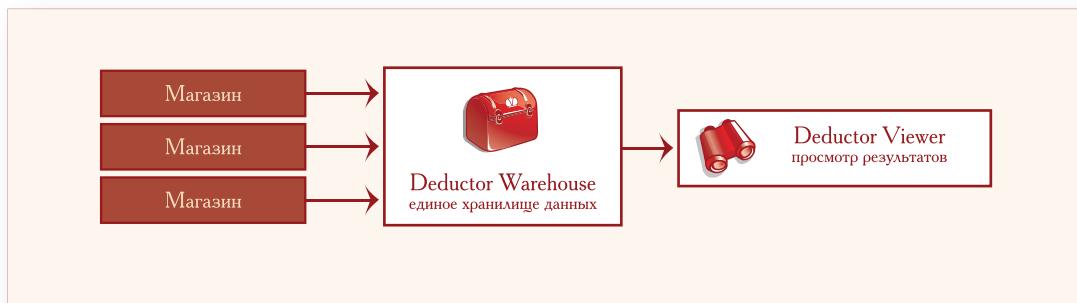
Торговая сеть «Дейли» ориентируется на формат «магазины возле дома» и состоит из множества относительно небольших магазинов.

Для работы с такой сложной территориально распределенной структурой необходима система поддержки принятия решений, консолидирующая данные из множества источников и дающая возможность увидеть картину в целом.

Аналитическая система, созданная в этой компании, представляет собой единое хранилище данных с подключенными к нему инструментами визуализации. Отлаженные процедуры загрузки данных позволяют все операции по консолидации данных проводить в автоматическом режиме в периоды минимальной нагрузки на оборудование. Для демонстрации результатов используется OLAP, диаграммы и другие способы отображения.

Консолидированные данные можно просмотреть в любом разрезе, «на лету» формулируя запросы с использованием понятных пользователю бизнес-терминов. Система самостоятельно генерирует запросы к базе данных и представляет нужные результаты в наиболее удобном для анализа виде.

Запуск аналитической системы позволил оптимизировать процесс управления, оперативно получать нужную информацию и принимать обоснованные решения. Кроме того, система является базой для других аналитических задач.



Образовательная программа

Образовательная программа

Реализованные на базе Deductor аналитические системы было бы невозможно создать без серьезной методической поддержки. Применение уже построенных сценариев не требует особого понимания методов анализа, но при построении моделей к анализу предъявляют серьезные требования.

Аналитик должен понимать методологию анализа, знать о задачах анализа и способах их решения, ведь цена его ошибки будет очень велика. Именно поэтому мы уделяем такое внимание методической поддержке.

В комплект поставки Deductor входит большой объем учебно-методических материалов: подробная справка, руководства аналитика и администратора, демонстрационный пример с подробным описанием, инструкции по работе с различными источниками данных, презентации, демонстрационные ролики и прочее. Разработаны практические кейсы по наиболее популярным задачам анализа с демонстрацией реальных сценариев обработки, описанием типовых проблем и способов их решения.

Deductor включен в учебную программу десятков высших учебных заведений в России, Украине, Белоруссии: Российская экономическая академия имени Г.В. Плеханова, Государственный университет управления, Санкт-Петербургский государственный университет, Нижегородский государственный технический университет, Белорусский государственный университет информатики и радиоэлектроники и множество других. Список ВУЗов, участвующих в образовательной программе BaseGroup Labs, постоянно расширяется, см. <http://www.basegroup.ru/edu/list/>.

Во второй половине 2007 года компанией открыт образовательный e-learning портал, посвященный технологиям анализа данных (<http://edu.basegroup.ru>) и рассчитанный на всех желающих получить или углубить свои знания в анализе данных. На портале размещено десятки модулей, включающих конспекты и практикумы по анализу данных и Data Mining. Теоретическая часть построена по системному принципу и охватывает темы: введение в анализ, консолидация, хранилища данных, ETL-технологии, очистка данных, визуализация, Data Mining и многое другое. Модульный принцип позволяет комплектовать курсы различной продолжительности и сложности.

Вместе с образовательной программой анонсирована бесплатная академическая версия Deductor Academic, на основе которой изучаются все практические аспекты технологий анализа данных.

Большая часть материалов образовательного портала находится в свободном доступе для самостоятельного изучения. Платный вариант предполагает использование преимуществ e-learning технологий: подбор индивидуальной программы обучения, общение с тьютором, консультации, самопроверки, проверка практических заданий, прохождение аттестаций на получение статуса сертифицированного аналитика и доступ к дополнительным материалам. Образовательная программа, e-learning технологии и Deductor Academic гарантируют получение качественных знаний в области анализа, отвечающих современным потребностям экономики, техники и бизнеса.

Распространение системы

Распространение системы

Deductor распространяется через партнерскую сеть. Партнеры осуществляют продажу системы, поддержку конечных пользователей, реализуют проекты на базе аналитической платформы.

Все партнеры — квалифицированные специалисты, которые проходят обучение и сертификацию, участвуют в реальных проектах.

Список партнеров BaseGroup Labs размещен на сайте — <http://www.basegroup.ru/partners/list/>, обратитесь к ближайшему партнеру для получения дополнительной информации.

Мы приложили много усилий, чтобы создать мощную, удобную и доступную по цене аналитическую систему. Воспользуйтесь Deductor и убедитесь, что современные аналитические технологии сделали глубокий анализ доступным.

Deductor позволяет совершенно по-новому взглянуть на данные, выжимая из них максимум ценной информации. Мы уверены, что грамотное использование Deductor принесет реальную пользу вашему бизнесу и повысит его конкурентоспособность.

Дополнительную информацию можно получить на сайте
www.basegroup.ru

Аналитическая платформа Deductor:

Единая интегрированная платформа

Все современные методики анализа бизнес-данных

Самообучающиеся алгоритмы построения моделей

Интегрированное хранилище данных

Высокая производительность

Простой пользовательский интерфейс

Курс дистанционного обучения

Дизайн Марина Агни, www.maredds.gallery.ru

Россия, 390046, Рязань

Введенская, 115, оф. 447

т./ф.: +7 (4912) 24-09-77

+7 (4912) 24-06-99

info@basegroup.ru

www.basegroup.ru